Argo real-time quality control intercomparison

Article

Accepted Version

Argo Real-Time Quality Control Intercomparison

R Wedd, *Australian Bureau of Meteorology*
M Stringer, *University of Reading, United Kingdom*
K Haines, *University of Reading, United Kingdom*

The real-time quality control (QC) methods applied to Argo profiling float data by the UK Met Office, the US Fleet Numerical Meteorology and Oceanography Centre, the Australian Bureau of Meteorology and the Coriolis Centre are compared and contrasted. Data are taken from the period 2007 to 2011 inclusive and real-time QC performance assessed with respect to Argo delayed-mode QC. An intercomparision of real-time QC techniques is performed using a common data set of profiles from 2010 and 2011. The real-time QC systems are found to have similar power in identifying faulty Argo profiles but to vary widely in the number of good profiles incorrectly rejected. The efficacy of individual QC tests are inferred from the results of the intercomparison. Techniques to increase QC performance are discussed.

INTRODUCTION

The accuracy of the initialized ocean state is key to forecasting short-range ocean conditions and seasonal global climates. Accurate initialization is heavily dependant on the quality and quantity of observational data. While remote sensing provides large quantities of information about the state of the ocean surface, the ocean sub-surface is not so easily observed. Prior to 2000 the primary methods of sub-surface observation were fixed buoy systems (TAO, TOGA)[1] and Ships of Opportunity (SOOP)[2]. Both of these methods have poor spatial sampling outside of the equatorial Pacific, and SOOP additionally have irregular temporal sampling. The Argo float project, implemented in 2000, provides high frequency observations with greater spatial uniformity.[3]

The standard operating procedure of an Argo float is to descend to drifting depth (usually 1000m) and follow ocean currents for a period of 8-10 days.[3] The float then descends to profiling depth (usually 2000m), and rises to the surface over a period of approximately 10 hours. Measurements of temperature, salinity and pressure are taken during the ascending phase. The float spends time at the surface (more than 10 hours for floats using the Argos communications systems, and ~30 minutes for floats using the Iridium system[4]), during which the recorded data is transmitted via satellite, and the next cycle begins. The measurements of a single ascent are called a 'profile', while each individual measurement is a 'level'. There are generally between 200 and 2000 levels in each profile, depending on which communication system is used.

Over 3600 Argo floats are currently recording oceanic profiles with a frequency of approximately 10 days.[5] This constitutes a large volume of important data. The highest standard of Argo quality control (QC) involves statistical analysis combined with direct scientific examination by experts. This process, known as delayed-mode QC (DMQC), is labour-intensive and cannot yet be implemented in time to be utilised in real-time operational applications. Fully automated QC is required to process the data in a timely fashion. This is known as real-time QC (RTQC).

A schematic demonstrating the flow of data from Argo floats to end users is shown in Fig 1.[6] National Argo Data Assembly Centres (DACs) collect the raw Argo profiles and implement standard RTQC tests set by the Argo Data Management group (ADM).[7] Profiles that pass the ADM RTQC are sent to the Global Telecommunications System (GTS) for download and use in operational oceanography. The DACs also send the full data, including RTQC flags, to the Global Argo Data Assembly centres (GDACs): Coriolis (CRS) in France and the US-GODAE in the US. Regional operational forecast centres download the Argo profiles from either or both of these sources and apply specialized RTQC procedures in addition to the ADM RTQC.

In 2002 the GODAE Ocean Data Quality Control Intercomparison Project was proposed.[8] RTQC data from GODAE members institutions was collected with the goal of an analytical intercomparison in the future. This work compares the RTQC methods utilized by members of GODAE OceanView[9], the successor of GODAE, using the collected RTQC data. Members must have available RTQC data in a readily readable format to be included. The included centres are: the United Kingdom Met Office (UKMO), the Australian Bureau of Meteorology (BoM), and the Fleet Numerical Meteorology and Oceanography Centre (FNMOC) in the USA. Data has also been included from the CRS GDAC. CRS is not an operational centre but does implement RTQC on Argo profiles for use in French Hydrographic Service (SHOM)[10] and French Research Institute for the Exploitation of the Sea (IFREMER)[11] operational products.

The benchmark for the intercomparison is the DMQC, which becomes available at a delay of several months to several years after the profile is recorded. The results of the DMQC process are the best assessment available and are considered here to be the 'truth'. The DMQC process involves expert operators checking temperature, salinity and pressure profiles manually for proper overall shape and consistency. The profile must be internally consistent across all variables and also consistent with readings from neighbouring floats. Historical data from the float is checked to assess sensor drift. Profiles are adjusted for detected sensor drift or offset, the profile error estimates are updated and the profile and levels are assigned QC flags.[7] In order to ensure sufficient DMQC data are available, only the years before 2012 have so far been considered. Data is taken from the five years from 2007-2012, as before 2007 CRS do not have data available. Profiles that have been adjusted by the DMQC are not included.

The assimilation of faulty profiles can be extremely detrimental to the quality of an analysis, especially if isolated in space and time. This is mitigated by the fact that profiles passing basic QC tests are less likely to have a catastrophic effect on an analysis: for example, a small drift in pressure will just add to systematic error, while a large temperature spike could produce spurious gravity waves that will detrimentally and persistently affect analysis accuracy. The first example could pass basic RTQC tests, while the second will be unlikely to do so. Removing catastrophically bad profiles should be the highest priority; the removal of slightly faulty profiles must be balanced against the systematic errors introduced by a lack of data. Though the Argo program provides many profiles for analysis, the spatial and temporal coverage is still sparse in comparison to the size of the global ocean: the ~3600 floats operating currently on 10 day cycles provide an average of 360 profiles daily, or approximately one profile per million square kilometres of ocean surface. The relative performance of the operational RTQC strategies must be assessed via the twin goals of removing the greatest number of faulty profiles while retaining as much as possible of this important observational resource.

DATA ANALYSIS

Profiles distributed via the GTS have the ADM RTQC applied before dissemination. Profiles and levels that fail are removed. UKMO and FNMOC take their data from this source. Profiles from the GDAC also have ADM RTQC applied but do not have any profiles or levels removed: the QC results are included in the profile meta-data. CRS use this data source. BoM employ a hybrid method: profiles are downloaded from both GDACs and the GTS. Duplicate profiles are removed by selecting the version with the greatest number of QC flags regardless of test results. The GDAC version of profiles is thus preferred, as the GTS system does not provide QC flag information.

The RTQC output flags differ from centre to centre, as do the data formats. All RTQC flags and the DMQC results are converted into a binary “good-profile/bad-profile” flag; the individual methods of achieving this are described in the Appendix. The comparison of the binary RTQC and DMQC outputs gives a contingency table of the form shown in Table 1. Four outcomes are possible: both agree the profile is bad (CB); both agree the profile is good (CG); RTQC flags the profile as good and DMQC as bad (FB); and RTQC flags the profile as bad and DMQC as good (FG). Two complimentary metrics are constructed from these outcomes; Recall (R) and Precision (P)[12]. Recall is a measure of success, and is defined as:

$$R = \frac{CB}{(CB+FB)} \quad [1]$$

Precision is a measure of accuracy, and is defined as:

$$P = \frac{CB}{(CB+FG)} \quad [2]$$

High R indicates the RTQC is identifying the majority of the DMQC-flagged bad profiles, and a high P indicates it is not removing many good profiles in the process. The results for the RTQC of each centre are shown in Table 2, 3, and 4, for temperature, salinity, and pressure profiles, respectively, for 2007-2011. FNMOC and UKMO do not apply QC to pressure measurements. There is a large range in the total size and the good/bad ratios of the initial data sets. Salinity profiles are the most likely to be assessed as faulty by the DMQC and pressure profiles the least likely. The centres using the GDAC system have larger initial data sets than the centres using the GTS and also have more faulty profiles.

The monthly averages of R and P for temperature, salinity and pressure for the five years studied are shown in Fig 2. The error bars represent the standard sampling error. For R this is defined as:

$$\frac{1}{\sqrt{T_B}} \quad [3]$$

where T_B is the total number of monthly profiles flagged as bad by the DMQC (FB+CB). For P it is defined as:

$$\sqrt{\left(\frac{1}{\sqrt{T_B}}\right)^2 + \left(\frac{1}{\sqrt{T_G}}\right)^2} \quad [4]$$

where T_G is the total number of monthly profiles flagged as good by the DMQC (FG+CG). The standard sampling error is an estimate of the uncertainty associated with measuring a generalized statistic using a random sample of size *n* and assumes the central limit theorem.[13] The metrics have been smoothed with a three month running average to reduce the noise from anomalous months and emphasise longer term trends in the results.

The results for temperature (Fig 2a and 2b) vary widely. BoM has a steadily increasing R with an average of approximately 0.7, and a steady P with the same average. The RTQC is identifying the majority of bad temperature profiles, while rejecting about half as many good profiles in the process. Data is available for the entire period with no breaks.

The FNMOC temperature results fluctuate between R values of approximately 0.2 to 0.4 and P values of approximately 0.1 to 0.2. The selection process is removing around a third of the bad profiles; a large number of good profiles are also being removed.

The UKMO temperature results show a distinct difference between the ASCII data of the 2007 to early 2008 period and the netCDF data after mid-2008 (see Appendix). The earlier results have lower R and higher P than the later data. This behaviour could be due to a broadening of the definition of a bad profile in the RTQC, rejecting greater numbers of both good and bad profiles. The gaps in the UKMO metrics indicate a lack of available data during those periods.

The CRS data set is the least temporally consistent of the included centres. The temperature RTQC has very good results in the first year. The data for 2007 is sparse, however, and reduces further in 2008. From March 2008 to December 2008 the metrics cannot be calculated as there are no profiles in the data that the DMQC has flagged as bad. During this time the RTQC is removing only DMQC-flagged good profiles. From 2009 P is only slightly lower than in 2007, but R drops by half. The reason for this drop is unknown. It is too large to be accounted for solely by sampling error. Data is again very sparse in the last half of 2009. In 2011 P is steady with a central value slightly below BoM, while R increases rapidly through the last half of 2011 until in February 2011 it is a similar value as BoM.

The RTQC results for salinity are similar to temperature for most of the centres. BoM identify fewer bad profiles and discard fewer good profiles (lower R, similar P). CRS show similar results for salinity as for temperature in 2007/8 but remove more bad and good salinity profiles in 2009 and 2010 (higher R, similar P). The increase in R from 2010 to 2011 that was seen in the CRS temperature results is also evident in salinity. FNMOC identify similar numbers of bad salinity profiles as temperature profiles but remove fewer good salinity profiles (similar R, higher P). Prior to 2009 the UKMO results are slightly higher in both R and P for salinity than for temperature. From 2009, the UKMO salinity RTQC has R and P approximately twice that for temperature. The UKMO salinity R results increase significantly in mid-2010. The reason for this is unknown.

BoM and CRS perform better for pressure than they do for salinity or temperature. The BoM pressure RTQC results vary more than those for salinity or temperature, but average slightly higher. The R values of CRS, though beginning low at the start of 2007, out-perform those of BoM after 2008, and the P values are extremely high. The two centres' results are very similar in 2011.

CRS and BoM both show a large dip in P for the pressure RTQC in the first half of 2010. This is not likely to be due to any change in the individual QC processes, as the drop is highly correlated between the centres in both phase and amplitude. Recall does not show any similar dip during this period. The monthly mean number of DMQC-identified bad profiles also remains stable for both centres. The number of profiles that are identified as bad by the RTQC increases dramatically, however, by 2-3 times the average of the surrounding months. The majority of these have 100% of their levels flagged as bad by the RTQC and 100% flagged as good by the DMQC. No meta-data for either centre indicate failure of any QC pressure test; the profiles seem to have been rejected based on other information.

A comparison between the data for the months of May, the lowest point of the dip, and October, two months after P recovers, shows the difference in the number of RTQC-identified bad profiles to be made up of Autonomous Profiling Explorer (APEX) type floats that have had their pressure profiles adjusted by the DMQC. This suggests the possibility that these profiles were rejected because the floats that recorded them were on a list of those affected by the Druck pressure sensor micro-leak issue[14]. These profiles could then have been cleared of any pressure drift by the DMQC. The micro-leak issue was discovered in early 2009, however, making it hard to reconcile the 2010 dip in the results.

The results shown in Fig 2 provide information regarding the temporal consistency of the RTQC processes and the performance of each centre over their individual data streams. The data sets analysed by the institutions are very different, however, as can be seen from the different data volumes in Tables 2, 3 and 4. A more meaningful comparison of the performance of each RTQC process requires homogeneity in the data set.

INTERCOMPARISON

Profiles that have undergone RTQC by each of the centres as well as the DMQC were isolated. The gaps in the Coriolis and UKMO early data and the small number of Coriolis profiles available during 2007 and the latter half of 2009 mean there are very few profiles with RTQC results from all centres before 2010. Only profiles from 2010 and 2011 were included, as the most stable, concurrent, and relevant period.

A possible bias was identified, introduced by profiles in which some levels have been removed by the Argo RTQC before dissemination on the GTS. The GTS profile in this case will have fewer levels than the GDAC or DMQC versions, which will lower the chance of GTS-based RTQC agreeing with the DMQC. To address this, all profiles in which enough of a disparity in levels exists between the DMQC and RTQC versions of the profile to possibly affect the outcome of a comparison for any centre were excluded from the intercomparison. This removes 5939 temperature profiles and 2658 salinity profiles. More detail is provided in the Appendix.

Isolating profiles that have undergone QC by all of the centres as well as the DMQC reduces the number of available profiles significantly due to differences between the centres' data sets. The number of profiles that meet the common requirement for temperature, salinity and pressure are shown in Table 5. Also shown are the number of DMQC-flagged bad profiles in the common data, the number of these bad profiles that are identified by each centres RTQC, and the number of DMQC-flagged good profiles that each centre rejects.

The results for each RTQC process are visualised in the Roebber diagrams[15] of Fig 3. This type of diagram utilises the geometric relationship between R, P, bias and critical success index (CSI) to display all four metrics simultaneously. R and P are defined in equations [1] and [2]. Bias is a measure of the relative frequency of selected and observed events, and is defined as:

$$Bias = \frac{CB+FG}{CB+FB} \quad [5]$$

where *CB, FG* and *FB* are defined in Table 1. Bias is indicated by radial dashed lines. CSI is a measure of accuracy when correctly identified good profiles are removed from consideration, and is defined as:

$$CSI = \frac{CB}{CB+FB+FG} \quad [6]$$

where *CB, FG* and *FB* are defined in Table 1. CSI is indicated by solid contour lines. CSI is not as relevant to this study as the other metrics due to the importance of correctly identifying good profiles. The errors in Fig 3 are the standard sampling error for R and P.

The performance of the RTQC processes show greater similarity over the common data set, especially in R. BoM and CRS change the most, while UKMO and FNMOC are close to their results in Fig 2, indicating that the GDAC-only profiles are being removed from consideration.

The R scores are not generally statistically distinguishable. All results in R for temperature profiles are within sampling error (Fig 3a). For salinity the only separations above sampling error are between FNMOC and the other centres (Fig 3b). The FNMOC salinity R result is the lowest of the centres. This can also be seen in the full data set for the years 2010-11 (Fig 2c). With the exception of FNMOC, the centres are more effective over salinity profiles than temperature profiles, generally identifying twice the proportion of bad profiles. The R results for pressure over the common data set are similar to those over the total data sets (Fig 3c). This is expected as both institutions use GDAC data. The BoM RTQC is narrowly more successful than CRS, though the separation is not significant. All centres identify between 17% and 19% of bad temperature profiles, between 38% and 41% of bad salinity profiles (excluding the FNMOC), and between 71% and 73% of pressure profiles.

The P scores show a larger spread in performance. UKMO and FNMOC remove significantly more good profiles than BoM or CRS for both temperature and salinity, with BoM removing the fewest in both variables (Figs 3a and 3b). All centres are more accurate for salinity profiles than temperature profiles. CRS have lower P scores than BoM for all three variables in the common data set, whereas in the full data CRS shows equal or greater P scores for 2010-11 (Figs 2b, d, f). The common data requirement is removing some profiles over which the CRS RTQC shows increased accuracy.

Similar R scores and different P scores lead to a spread in bias for both temperature and salinity. The two outliers are BoM and FNMOC: BoM remove 65% fewer total temperature profiles than the DMQC (bias of 0.35), while FNMOC remove 66% more (bias of 1.66). The lower P score of FNMOC indicates that the extra profiles removed are flagged as good by the DMQC. For salinity UKMO have the highest bias (1.18) while BoM are again the lowest (0.52).

DISCUSSION

The RTQC criteria employed by each of the examined centres are shown in Table 6. The sources of the information are listed in the table caption. Details of the FNMOC RTQC test criteria were sparse in the available documentation. There is a common general strategy across the centres. Physical value tests are applied to the date and time of profile recording, and the position of the profile is required to be physical and within a defined ocean space. All institutes except BoM set a maximum allowed drift speed. BoM and FNMOC require temperature and salinity values to be within single globally-applied physical value test, while CRS applies three geographically dependant physical value tests to profiles before 2011 and adds tests for the north western shelves and Arctic sea regions for profiles after 2011. UKMO

do not apply a global range test but do apply a tropical waters test, rejecting any level above 1000m that measures below 1°C.

All of the centres apply monotonicity/inversion tests to density. Salinity and temperature for levels which fail are flagged as bad. UKMO also apply a density spike test. FNMOC and BoM test depth data: the FNMOC rejecting levels with duplicate depth and enforcing monotonicity; BoM applying a global bathymetry test, rejecting levels with greater depth than that interpolated from a 2" gridded bathymetry product with enhanced resolution around Australia.

The main differences between the temperature RTQC processes lie in the sophistication of the gradient, spike and density tests, and the different background and depth tests. The CRS gradient test compares the value at a level with the average value of the two adjacent levels; the maximum permitted deviation has two values, split at 500dB. BoM use a similar test, but weight the average value of the two adjacent levels according to their relative distance from the level being tested. UKMO do not employ a gradient test; they implement a more stringent spike test than BoM or CRS, however. FNMOC apply a simple maximum gradient requirement, and compare with climatological gradients. FNMOC also include temperature spike and inversion tests; no details of these were available. All centres require increasing density between levels; FNMOC and UKMO have minimum density change requirements while BoM and CRS just require an increase. UKMO additionally apply a density spike test.

BoM, UKMO and CRS compare the profile to a background state; hybrid CARS/WOA05 climatology[16 17 18] for BoM, a Bayes theorem-based[19] check against a one-day forecast for UKMO, and an objective analysis using climatology derived from WOA98[20] as the background for CRS. In 2010 CRS updated their objective analysis process; this combined with the added physical value tests could explain the upswing in R seen in Figs 2a and 2c. FNMOC do not apply a background check.

The gradient tests seem to be reducing the number of incorrectly rejected profiles (FG) for BoM and CRS over UKMO. The FNMOC gradient test is simple and might not provide the same level of discrimination as the CRS and BoM tests. The weighting of the gradient test could be a source of added accuracy for BoM over CRS; the difference could also be attributed to the different climatological background checks, the added depth check or the lack of speed and sensor drift tests in the BoM RTQC.

The test criteria for salinity are very similar to those for temperature. FNMOC alone of the centres studied has lower R over salinity profiles than over temperature profiles. The poor result is difficult to interpret due to the sparse details of the FNMOC RTQC process in the available documentation. The difference in salinity criteria between the other three centres again lie in the gradient, climatology and density inversion tests. The UKMO show the largest increase in R between salinity and temperature and identify the greatest number of faulty salinity profiles, though the result is within sampling error of BoM and CRS. BoM's method again has the highest P score, though the gap between CRS is narrowed.

Pressure is tested less stringently than salinity or temperature by both BoM and CRS: BoM apply only physical value and monotonicity/inversion tests; CRS also apply a deepest pressure test. The small performance difference between BoM and CRS pressure QC can be attributed to the different ranges used in the physical value test, a negative impact from the CRS deepest pressure test, the different background checks, or the added bathymetry test of the BoM RTQC. The difference in performance is not statistically significant in any metric.

The performance of the RTQC methods over identical profiles is not the only consideration in determining the preferred treatment of Argo data: the choice of data source is also relevant. Tables 7, 8 and 9 show the data for 2010-2011 without the common QC assessment requirement but with the removal of the level-based bias for temperature, salinity and pressure, respectively. The initial data sets of the centres vary widely. Despite taking profiles from the same source, the FNMOC data set is approximately 20% larger than the UKMO. The difference is believed to be made up of US Navy Argo profiles that are not shared in real time due to security concerns. There is also a difference in size of the initial data sets of those institutions sourcing data from the GDACs. Though the BoM's hybrid data collection method is designed to collect all profiles available from all sources, CRS have ~1000 more profiles. This is believed to be the contribution of the French Navy Argo floats, which are also not available in real time. The removal of these profiles

from the common data set could explain the comparative drop in P results for CRS between the full and common data sets.

Tables 7 and 8 show that although the BoM's RTQC method gives the best result over the common data, BoM do not have the fewest bad profiles in their final data sets. This is due to the different data streams. The GTS provides UKMO with smaller and cleaner initial data sets than the BoM's GDAC/GTS hybrid method. The UKMO final data sets also have fewer good profiles and bad profiles than the BoM's. For temperature, BoM's method results in 216 more bad profiles and 17 313 more good profiles. For salinity, the BoM have 423 more bad profiles and 12 878 more good profiles.

The optimal balance between the removing bad profiles and retaining good ones is system dependant. A system with an accurate ocean model will have less need for corrective data and should seek to remove as many error-inducing profiles as possible. A system with a long assimilation period will have a lower likelihood of erroneous data being isolated in space and time, and could thus include more bad profiles and rely on the mitigating effects of concurrent good data. The choice of a smaller, more accurate data set via the GTS or a larger, more contaminated one via the GDACS must be made by the individual user.

SUMMARY AND CONCLUSIONS

The performance of the real-time Argo profile QC methods used at the UK Met Office, the Australian Bureau of Meteorology, the Fleet Numerical Meteorology and Oceanography Centre and the Coriolis centre were assessed and compared. The RTQC output of the centres was obtained for the years 2007 to 2011 inclusive. The data were brought into a common binary good/bad-profile format and compared to the DMQC results. The data sets of some of the institutions were found to be temporally irregular due to changes in the RTQC criteria, recording methods and/or gaps in the uploaded data. An intercomparison was performed using those profiles recorded in 2010 and 2011 which had undergone RTQC by all institutions and the DMQC.

The RTQC techniques were found to identify similar numbers of faulty profiles; BoM slightly more temperature and pressure profiles and UKMO slightly more salinity profiles. The differences between the systems were not generally statistically significant given the data available. That the FNMOC identifies fewer faulty salinity profiles than any other centre can be stated confidently, but no other differences are significant. The number of good profiles rejected in the RTQC process was more system-dependant. The BoM's RTQC was found to remove significantly fewer good temperature and salinity profiles than the other RTQC techniques. CRS and BoM remove a similar number of good and bad pressure profiles.

The GTS distribution stream removes profiles that fail the Argo RTQC before dissemination, while the GDACs supply all profiles reported by Argo buoys along with the flags from the Argo RTQC. This results in very different pre-RTQC data sets for each centre, and the best performing RTQC system does not necessarily provide the cleanest final set of profiles. FNMOC and CRS have access to military-operated floats that are not generally disseminated in real-time and their data sources cannot be compared to other centres. The pre- and post-RTQC data of UKMO and BoM were compared for 2010-2011. The UKMO RTQC using GTS data has fewer faulty temperature and salinity profiles than the BoM system using their hybrid GTS/GDAC data source. It also results in many fewer good profiles in the final data set. Whether the removal of the extra bad profiles is worth the removal of the good profiles is dependant on the model and assimilation systems being used, and the choice must be left to the individual user.

The accuracy of operational ocean forecasting is directly related to the quality of the Argo data stream. Improving the Argo RTQC processes of operational centres will also provide material benefit to seasonal and decadal forecasts via better initialisation. The current RTQC techniques remove many faulty Argo profiles, but there is scope for improvement. One possibility is to investigate the automation of DMQC tests, for example: the parametrisation of expected profile shapes; drift analysis using previous profiles from the same float; or algorithmic temperature/salinity/pressure profile consistency checks. Statistical methods such as these can perform poorly in the presence of ocean features such as eddies, fronts, and water mass boundaries, however, and care would have to be taken to avoid discarding profiles with extreme but valid attributes. Another possibility is the creation of a super-RTQC

assessment based on combining the results of the individual centres' RTQCs using classical statistical methods or machine learning techniques. The differences in the RTQC techniques could perhaps be leveraged for greater discriminatory power. This would need to be a centralised process, and the classification of profiles in real-time would rely on prompt and consistent uploading of RTQC results from operational centres.

The Argo program continues to provide operational centres and researchers with high quality sub-surface ocean observations. Operational RTQC systems should be subject to continual analysis and improvement as the Argo data base grows. Updating the results shown here as more DMQC data becomes available will lower the statistical errors and highlight the differences between the RTQC systems, helping to inform the improvement of current RTQC systems and the design of new ones.

AKNOWLEDGEMENTS

The authors would like to thank Alastair Gemmel of the University of Reading and James Cummings of the United States Naval Research Laboratory for their previous work on the subject of Argo quality control. The feedback and suggestions from Gary Brassington of the BoM, and Simon Good and Matthew Martin of the UKMO were instrumental in achieving the results described here.

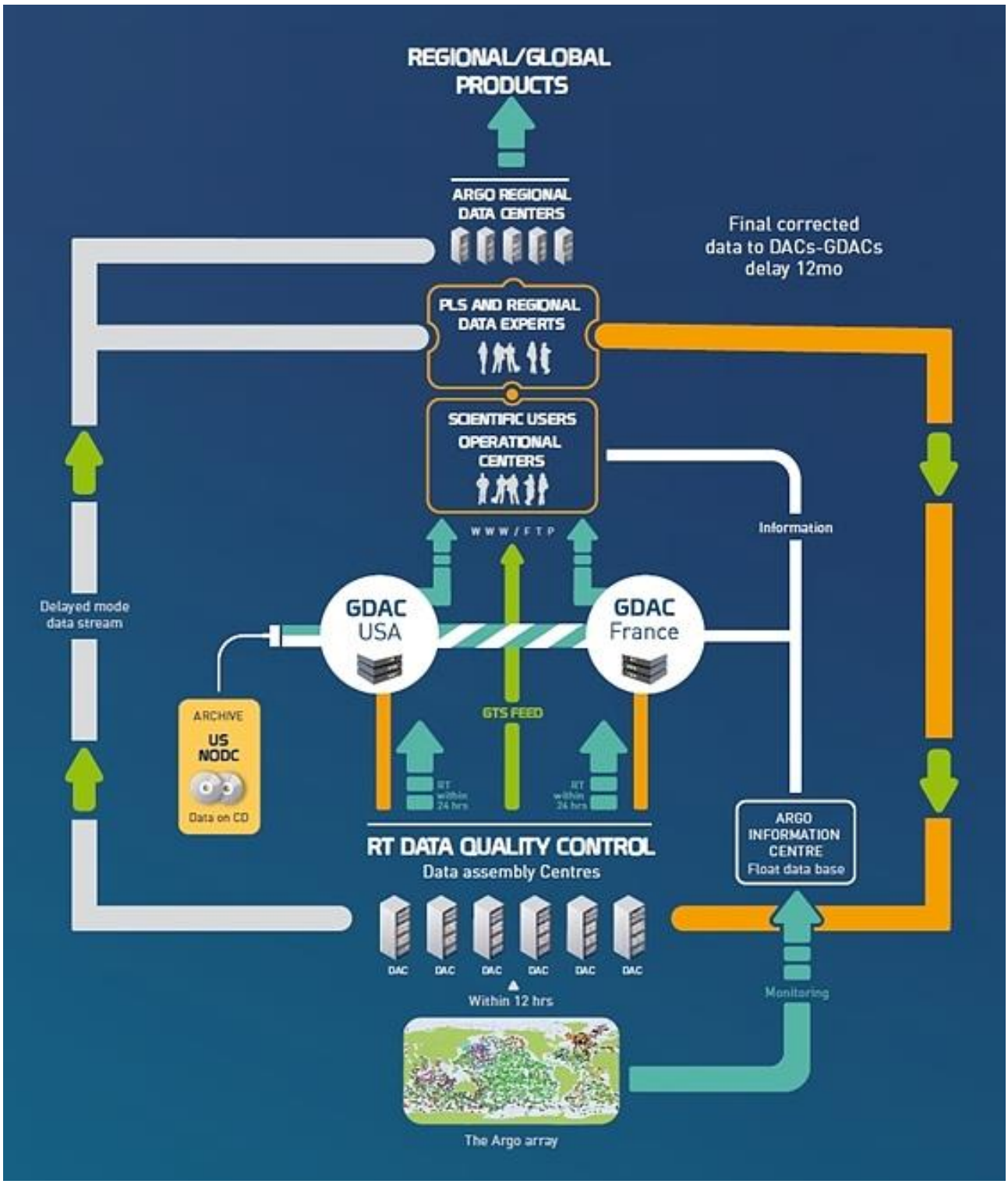

Fig 1: Schematic of data flow from the Argo float array through the DACs to operational centres via the GTS and GDACs.[6]

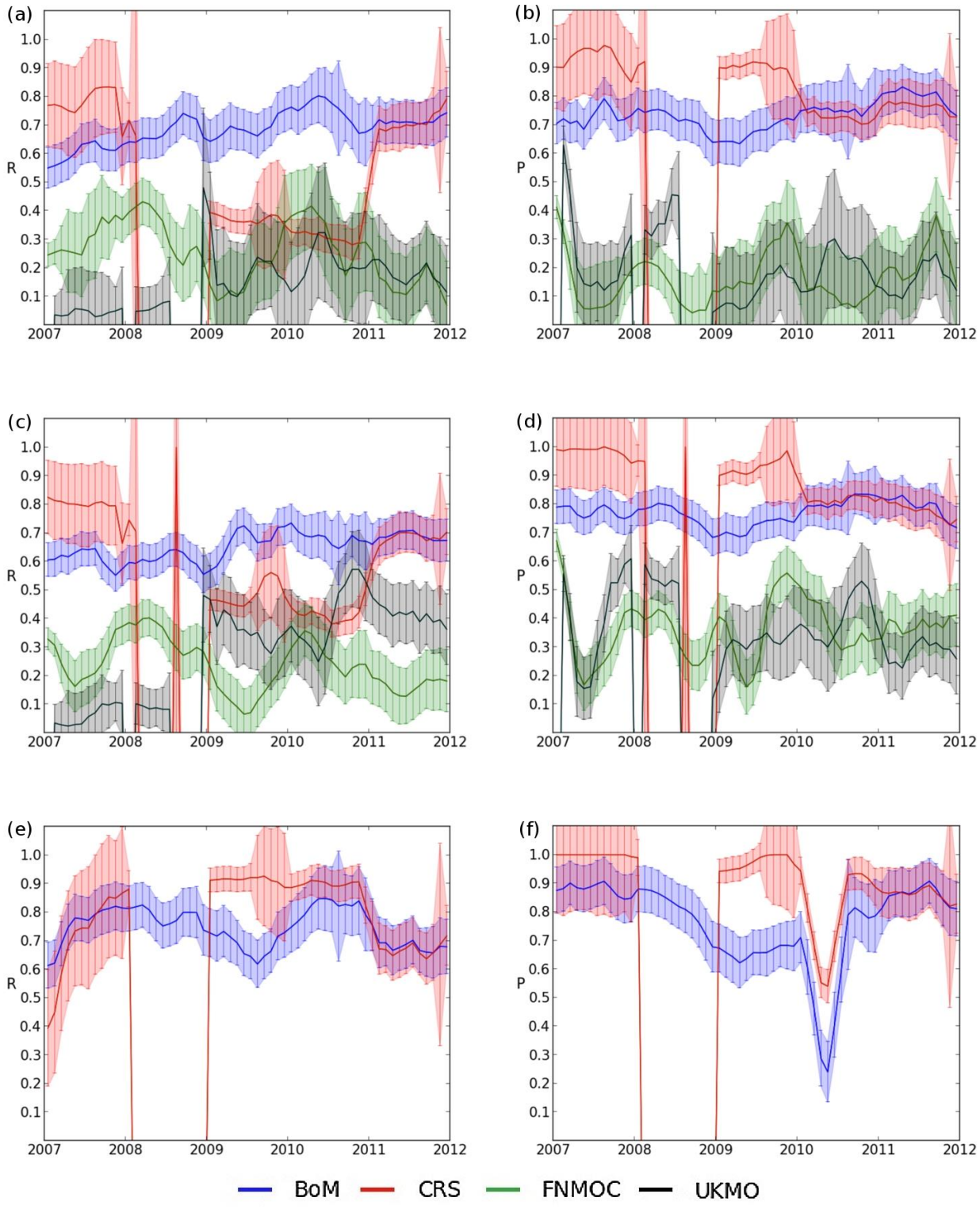

Fig 2: Recall and Precision for RTQC processes. R (P) for temperature is shown in 1a (1b), for salinity in 1c (1d), and for pressure in 1e (1f). Error bars are the standard sampling error. A three month running average has been applied. High scores are desirable for both R and P.

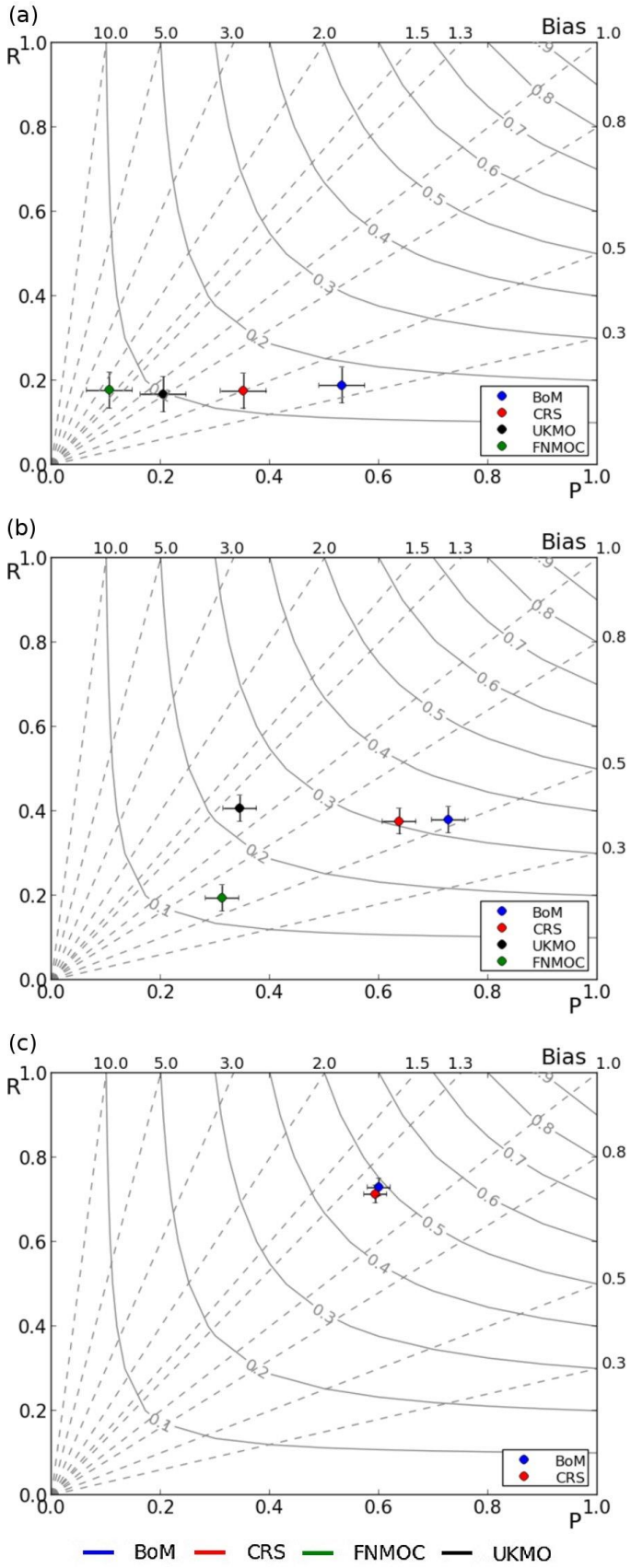

Fig 3: Roebber diagrams of temperature (a), salinity (b) and pressure (c) RTQC results in 2010 and 2011. Error bars are the standard sampling error. Bias is shown in radial dashed lines. CSI is shown in solid contours. Ideal performance lies in the upper-right of the diagram where all metrics approach 1.0.

		DMQC	
		Good	Bad
RTQC	Good	CG	FB
	Bad	FG	CB

Table 1: Contingency table defining the possible outcomes of a comparison between binary DMQC and RTQC. CG denotes 'correct good', FG 'false good', FB 'false bad', and CB 'correct bad'.

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	182274	11949	170325	5232	1185	175857	6717	169140	0.44	0.82
BoM	365241	10775	354466	7484	2603	355154	3291	351863	0.69	0.74
UKMO	247307	3037	244270	338	3805	243164	2699	240465	0.11	0.08
FNMOC	356797	5090	351707	1285	7653	347859	3805	344054	0.25	0.14

Table 2: RTQC results for temperature from 2007-2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	182221	14426	167795	7441	1222	173558	6985	166573	0.52	0.86
BoM	364894	15593	349301	10298	2872	351724	5295	346429	0.66	0.78
UKMO	246624	5284	241340	1312	4668	240644	3972	88352	0.25	0.22
FNMOC	356808	9598	347210	2290	3796	350722	7308	343414	0.24	0.38

Table 3: RTQC results for salinity from 2007-2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	182232	10500	171732	9134	1622	171476	1366	170110	0.87	0.85
BoM	365252	8653	356599	6529	2410	356313	2124	354189	0.75	0.73

Table 4: RTQC results for pressure from 2007-2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

	Initial Profiles	DMQC Bad	CB				FG			
			CRS	BoM	UKMO	FNMOC	CRS	BoM	UKMO	FNMOC
T	61641	568	99	107	95	100	182	94	368	845
S	62876	1073	403	407	436	208	229	153	826	457
P	89120	2272	1620	1655	--	--	1109	1105	--	--

Table 5: RTQC results for the common data set from 2010-2011. Listed are the total initial profiles, the number of DMQC-flagged bad profiles, the DMQC-flagged bad profiles that each RTQC rejects (CB) and DMQC-flagged good profiles that each RTQC rejects (FG).

	CRS	FNMOC	BoM	UKMO
Pressure	**Physical value test** Level fails if P<-5dBar **Monotonicity test** (k+1) fails if P(k+1)<=P(k) **Deepest pressure test** Level fails if pressure is greater than 10% higher than the deepest pressure	None.	**Physical value test** Level fails if it does not satisfy: 0<=P<=6500 dBar **Monotonicity test** (k+1) fails if P(k+1)<=P(k)	None.
Temperature	**Range test** Level fails if it does not satisfy: -2.5°C<T<40°C 21.7°C<T<40°C for Red Sea 10°C<T 40°C for Mediterranean plus, for post-2010 profiles: -2°C<T<24°C for North Western Shelves -2°C<T<30°C for South West Shelves -1.92°C<T<25°C for Arctic Sea **Gradient test** X = |T(k)-(T(k-1)+T(k+1))/2| k fails if X>9°C and P<500dB k fails if X>3°C and P>=500dB **Spike test** X = |T(k)-(T(k-1)+T(k+1))/2| -|(T(k-1)- T(k+1))/2| k fails if X>6°C and P<500dB k fails if X>2°C and P>=500dB **Digit rollover test** k fails if |T(k)-T(k-1)| or |T(k)-T(k+1)|>10°C **Stuck value test** Profile fails if all profile values are identical	**Global range test** Level fails if it does not satisfy: -2.5°C<=T<=42°C **Spike test** no details **Inversion test** no details **Gradient test** Level fails if gradient>0.2°C/m and 4σ from climatological gradient **Land-Sea Boundary test** no details	**Global range test** Level fails if it does not satisfy: -2°<=T<=40°C **Missing value test** **Gradient test** k fails if GRAD(k)>10°C, for h<=500 m k fails if GRAD(k)>5°C, for h>500 m where GRAD(k) = T(k)-(alpha1*T(k-1)-alpha2*T(k+1)) alpha1 = (h(k+1)-h(k))/(h(k+1)-h(k-1)) and alpha2 = (h(k)-h(k-1))/(h(k+1)-h(k-1)) where h(i) are depth values. **Spike tests** As UKMO except: Ttol=5°C if depth<=500m, Ttol=2.5°C if depth>500m	**Constant Value test** If over 90% of T levels that cover at least 100m read identical, T profile fails **Tropical waters test** If depth<1000m and T<=1°C reject level **Spike tests** **1)** If either |DT(k-1)|>Ttol or |D(Tk)|>Ttol, and |DT(k-1)+DT(k)|<0.5Ttol then T(k-1) is rejected as a spike. **2)** If |DT(k-1)|>0.5Ttol or |DT(k)|>0.5Ttol, and there exists k, where GRAD(T(k))>0.05°C/m, and |DT(k-1)+DT(k)|<0.25*|DT(k-1)-DT(k)| then T(k-1) is rejected as a spike. **3)** If |DT(k-1)|>Ttol, and |T(k-1)|>0.5Ttol(interpolated T(k-1) and T(k)), and 0<DT(k-1) or DT(k-1)<-3×TTol (d<250m), then T(k-2) and T(k-1) are flagged as suspect. Where DT(k)=T(k)-T(k-1), and Ttol=5°C (<300m), 2.5°C(<500m), 2.0°C(<600m), 1.5°C(>600m). If within 20° of the equator then 200m=300m and 300m=400m. Ttol is linearly interpolated from 0m to 300m(400m), step function after that.
Salinity	**Global range test** Level fails if it does not satisfy: 2<S<41 psu 2<S<41 psu for Red Sea 2<S<40 psu for Mediterranean plus, for post-2010 profiles: 0<S<37 psu for North Western Shelves 0<S<38 psu for South West Shelves 2<S<40 psu for Arctic Sea **Gradient test** X=|S(k)-(S(k-1)+S(k+1))/2| k fails if X>1.5 psu and P<500dB k fails if X>0.5 psu and P>=500dB **Spike test** X = |S(k)-(S(k-1)+S(k+1))/2] -|(S(k-1)+S(k+1))/2| k fails if X>0.9 psu and P<500dB k fails if X>0.3 psu and P>=500dB **Digit rollover test** k fails if |S(k) - S(k-1)| or |S(k) - S(k+1)|>5 psu **Stuck value test** Profile fails if all profile values are identical	**Physical value test** Level fails if it does not satisfy: 0<=S<=42 psu	**Global range test** Level fails if it does not satisfy: 0<=S<=39 psu **Missing value test** **Gradient test** similar to T but with limits of GRAD(k)<1 psu, h<=500m GRAD(k)<1 psu, h>500m **Spike test** As UKMO except: Stol=1 psu if depth<=500m Stol=0.2 psu if depth>500m	**Temperature Profile test** if >50% of the T profile is bad, the S profile is rejected. **Constant Value test** If 70% or more of S levels over at least 50m are identical, S profile fails **Spike test** Similar to T but only tests 1 and 3, and without the 0<DS(k-1) or DS(k-1)<-3×TSol (d<250m) condition in 3. Stol=1 psu (<300m), 0.2 psu(>300m). If within 20° of the equator then 300m=400m. Stol is linearly interpolated from 0m to 300m(400m), step function after that. If a T spike is detected the corresponding S value is automatically rejected. If >4 T spikes then both T and S profiles are rejected.

	CRS	FNMOC	BoM	UKMO
Density	**Inversion test** Calc density (D) from T and S T and S level k fails if D(k)>D(k+1) or D(k)<D(k-1)	**Inversion test** Level k fails if D(k)-D(k-1)<-0.025 kg.m^-3	**Monotonicity test** If T and S tests are passed; Level fails if D(k)<=D(k+1)	**Monotonicity test** Dρ(k)=ρ(Θ(k),S(k),P(k))-ρ(Θ(k-1),S(k-1), P(k)). If Dρ(k)>-0.03 kgm^-3 then T and S fail **Density spike test** If |Dρ(k-1)+Dρ(k)|>0.25*|Dρ(k-1)-Dρ(k)| then fail T and S at k-1. If both tests fail then T and S at k and k-1 fail. If a profile has 2 or more inversions the profile is discarded.
Bathymetry / Depth	None.	**Duplicate depth test** no details **Monotonicity test** no details	Level fails if deeper than interpolated 2 minute global bathymetry formed from 2" ETOPO2v2[21] and 1km Geosciences Australia[22] products.	None.
Date	Profile fails if it does not satisfy: Year>1997 1<=Month<=12 Day exists in Month 0<=Hour<=23 0<=Minute<=59	Profile fails if it does not satisfy: 1<=Month<=12 Day exits in Month 0<=Hour<=23 0<=Minute<=59 0<=Second<=59 Observation time must be older than the receipt time at the centre	As CRS	As CRS
Position	Profile fails if it does not satisfy: -90<=Latitude<=90 -180<=Longitude<=180 Must be in ocean (ETOPO5)	Profile fails if it does not satisfy: -90<=Latitude<=90 -180<=Longitude<=180 Must be in ocean	As CRS	As CRS
Speed	If drift speed > 3m/s then flag time, position and/or float number as wrong.	Speed < 2m/s	**None.**	Speed(K)= (Dist(K)-0.5DistRes)/MAX(DTime,TimeRes) where DistRes=20km and TimeRes=600s. If speed>2m/s, or >1.6m/s and there is a kink in the track, then a series of checks are run to determine which position correct. If a buoy has >50% of it's profile positions rejected the buoy is removed.
Background	Objective analysis check using climatology derived from WOA98[20] as background.	None.	T and S mean within 5σ of CARS[16] climatology within 71°S - 26°N for all longitudes and WOA[17 18] in all other regions.	Bayesian background probability check. A 1 day ocean forecast is used as background.

Table 6: RTQC test criteria for CRS[23 24 25], FNMOC[26 27], BoM[28] and UKMO[29].

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	100331	6065	94266	2479	827	97025	3586	93439	0.41	0.75
BoM	97677	3210	94467	2351	665	94661	859	93802	0.73	0.78
UKMO	81968	789	81179	146	516	81306	643	80663	0.19	0.22
FNMOC	100345	991	99354	197	1529	98619	794	97825	0.20	0.11

Table 7: RTQC results for temperature from 2010 and 2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	101904	7231	94673	3527	884	97493	3704	93789	0.49	0.80
BoM	99360	4417	94943	3093	795	95472	1324	94148	0.70	0.80
UKMO	84201	1567	82634	666	1364	82171	901	81270	0.43	0.33
FNMOC	102443	2201	100242	419	779	101245	1782	99463	0.19	0.35

Table 8: RTQC results for salinity from 2010 and 2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

Centre	Initial Profiles	DMQC Bad	DMQC Good	RTQC CB	RTQC FG	Final Profiles	RTQC FB	RTQC CG	R	P
CRS	103133	5492	97641	4618	1392	97123	874	96249	0.84	0.77
BoM	100706	2630	98076	1938	1235	97533	692	96841	0.74	0.61

Table 9: RTQC results for pressure from 2010 and 2011 for the full independent data from each institution. Listed are the total initial data sets of profiles that have results for both RTQC and DMQC, the number of DMQC-flagged good and bad profiles in the initial data sets, the number of DMQC-flagged good and bad profiles that the RTQC rejects, the total post-RTQC final data set, the number of DMQC-flagged good and bad profiles in the final data set, and the calculated values of R and P.

REFERENCES

1. Hayes SP, Mangum LJ, PiCaut J, Sumi A, Takeuchi K. 1991. *TOGA-TAO: A Moored Array for Real-time Measurements in the Tropical Pacific Ocean*, Bulletin of the American Meteorological Society **72(3)**: 339-347.

2. Goni G, Roemmich D, *et al*. 2010. *The ship of opportunity program*. Proceedings of OceanObs 9.

3. Roemmich D, Boebel O, Freeland H, King B, LeTraon P, Molinari R, Brechner Owens W, Riser S, Send U, Takeuchi K, Wijffels S. 1998. *On The Design and Implementation of Argo: an initial plan for a global array of profiling floats*, International CLIVAR Project Office Report 21.

4. Roemmich D, Johnson GC, Riser S, Davis R, Gilson J, Owens WB, Garzoli SL, Schmid C, Ignaszewski M. 2009. *The Argo Program: Observing the global ocean with profiling floats*. Oceanography **22(2)**: 34–43.

5. Current status of the Argo fleet taken at 8/7/2014. Available at www.argo.uscd.edu

6. Copywrite Euro-Argo. http://www.euro-argo.eu/

7. Wong A, Keeley R, Carval T, *et al*. 2013. *Argo quality control manual. Version 2.9.*

8. Cummings J, Brassington G, Keeley R, Martin M, Carval T. 2010. *GODAE ocean data quality control intercomparison project*. Proceedings of OceanObs 09: Sustained Ocean Observations and Information for Society.

9. Le Traon PY, Bell M, Dombrowsky E, Schiller A, Wilmer-Becker K, *et al*. 2010. *GODAE OceanView: from an experiment towards a long-term ocean analysis and forecasting international program*. Proceedings of OceanObs 09. GODAE OceanView collaboration homepage www.godae-oceanview.org

10. http://www.shom.fr

11. http://wwz.ifremer.fr

12. Van Rijsbergen, CJ. 1979. *Information Retrieval* (2nd edit.) Butterworths. London, UK.

13. Isserlis L. 1918. *On the Value of a Mean as Calculated from a Sample*, Journal of the Royal Statistical Society, **81(1)**: 75-81.

14. Barker PM, Dunn JR, Domingues CM, Wijffels SE. 2011. *Pressure Sensor Drifts in Argo and Their Impacts*. J. Atmos. Oceanic Technol., **28(1)**: 1036–1049.

15. Roebber PJ. 2009. *Visualizing multiple measures of forecast quality*, Weather and Forecasting **24(2)**: 601-608.

16. Ridgway KR, Dunn JR, Wilkin JL. 2002. *Ocean interpolation by four-dimensional least squares -Application to the waters around Australia*, J. Atmos. Ocean. Tech., **19(9)**: 1357-1375.

17. Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE. 2006. *World Ocean Atlas 2005, Volume 1: Temperature*. S. Levitus, Ed. NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C.: 182.

18. Antonov JI, Locarnini RA, Boyer TP, Mishonov AV, Garcia HE. 2006. *World Ocean Atlas 2005, Volume 2: Salinity*. S. Levitus, Ed. NOAA Atlas NESDIS 62, U.S. Government Printing Office, Washington, D.C.: 182.

19. Lorenc AC, Hammon O. 1988. *Objective quality control of observations using Bayesian methods. Theory, and a practical implementation*, Q.J.R. Meteorol. Soc., **114**: 515–543.

20. US Department of Commerce, National Oceanic and Atmospheric Administration. 1998. *World ocean atlas 1998*.

21. US Department of Commerce, National Oceanic and Atmospheric Administration. 2006. *2-minute Gridded Global Relief Data (ETOPO2v2)*.

22. Webster MA, Petkovic P. 2005. *Australian bathymetry and topography grid, June 2005.* Geoscience Australia Record, **12:** 30.
23. Wong A, Keeley R, Carval T. 2009. *"Argo quality control manual. Version 2.32"*

24. Pouliquen S, *et al.* 2011. *Recommendations for in-situ data Near Real Time Quality Control*, EuroGOOS.

25. Gaillard F, Autret M, Thierry V, Galaup P, Coatanoan C, Loubrieu T. 2009 *Quality Control of Large Argo Datasets*, J. Atmos. Oceanic Technol., **26**: 337–351.

26. Cummings JA. 2006. *Operational multivariate ocean data assimilation*, Q. J. R. Meteorol. Soc., **131**: 3583–3604.

27. Cummings JA. 2010. *Ocean Data Quality Control*, Operational Oceanography in the 12st Century, Springer: 91-121.

28. BLUElink team. *BLUElink Real-Time Quality Control File Format*, Australian Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organisation, unpublished.

29. Ingleby B, Huddleston M. 2007. *Quality control of ocean temperature and salinity profiles — Historical and real-time data*, Journal of Marine Systems, **65(1–4)**: 158–175.

APPENDIX

The RTQC output flags differ from centre to centre. The Argo DMQC flags follow the form of table 2a in the Argo Data Management User Manual[4] (see Table A1), in which the flags run from A-F, based on the percentage of levels in the profile that were assessed as 'good' data. BoM and CRS also follow table 2a. UKMO RTQC flags follow the form of table 2 in the same publication (see Table A2), and FNMOC use their own probability-based method. Two methods of bringing the flags into a common form for comparison were examined. The first converts all QC systems into a binary "good-profile/bad-profile" flag; the individual methods of achieving this are described below. This is the form used in the majority of the analysis. The second method involves converting all RTQC flags into the form of Table 2a by calculating the proportion of levels in each profile that are rated 'good' according to the centre's RTQC. No appreciable difference was found between the methods. The first method is preferred due to the possibility of some profile-based RTQC tests being excluded from the level-based RTQC results.

Argo Delayed-mode:

The Argo delayed-mode data was obtained from the US GODAE public server (http://www.usgodae.org/pub/outgoing/argo/geo/). The profile-based QC flags follow the form of Table 2a. All profiles with greater than or equal to 50% of levels rated as 'good' (flag A, B or C) are considered a 'good' profile for the purposes of this intercomparison, and all profiles with less than 50% of levels rated as 'good' are considered 'bad'. Profiles that have been adjusted by the delayed mode QC operators are excluded.

UKMO:

The UKMO's data is in the form of both ASCII and netCDF files that are uploaded to the GODAE severs daily (http://www.usgodae.org/pub/incoming/godae_qc/). The ASCII data runs from 10/4/2006 to 22/7/2008, and the netCDF data runs from 17/12/2008 to the present. Data for the gap of roughly five months were unavailable. The profiles are flagged according to Table 2. All profiles and levels rated as 'Good' or 'Probably Good' (flag 1 or 2) are considered to be 'good'. All other flags are considered 'bad'.

BoM:

The BoM has provided daily netCDF files spanning the period 1/1/2005 to 19/6/2011 to the GODAE servers. Data for the rest of 2011 was obtained from the Bureau directly. The level-based RTQC is flagged according to Table 2, and the profile-based RTQC according to Table 2a. As with the DMQC data, profiles with greater than 50% of levels rated as 'good' (flag A, B or C) are considered 'good'.

CRS:

CRS uses the same format as BoM and is treated in the same fashion. Data is available from 23/7/2009 to the present on the GODAE servers, but the data for 2009 is very sparse. Further data was obtained from the MyOcean ftp server (http://www.mycean.eu/) for the years 2007-2008. While files were available from the beginning of 2007 to the end of 2009, the data is very sparse at all times except the first half of 2009.

FNMOC:

The raw data files from FNMOC were unavailable; data that were processed into float specific files were used instead. These are available on the GODAE public server (http://www.usgodae.org/pub/outgoing/godae_qc/) from 2004 to 2011. FNMOC assigns profiles a probability value of being 'bad' from 0 to 100. Profiles with probabilities of less than 96 are considered to be 'good'. Levels with probabilities of less than 100 are considered to be 'good'.

n	Meaning
" "	No QC performed
A	***N*** = 100%; All profile levels contain good data.
B	75% <= ***N*** < 100%
C	50% <= ***N*** < 75%
D	25% <= ***N*** < 50%
E	0% < ***N*** < 25%
F	***N*** = 0%; No profile levels have good data.

Table A1: Argo Data Management User Manual Table 2a.

n	Meaning	Real-time comment	Delayed-mode comment
0	No QC was performed	No QC was performed	No QC was performed
1	Good data	All Argo real-time QC tests passed.	The adjusted value is statistically consistent and a statistical error estimate is supplied.
2	Probably good data	Probably good data	Probably good data
3	Probably bad data that are potentially correctable	Test 15 or Test 16 or Test 17 failed and all other real-time QC tests passed. These data are not to be used without scientific correction. A flag '3' may be assigned by an operator during additional visual QC for bad data that may be corrected in delayed-mode.	An adjustment has been applied, but the value may still be bad.
4	Bad data	Data have failed one or more of the real-time QC tests, excluding Test 16. A flag '4' may be assigned by an operator during additional visual QC for bad data that are uncorrectable.	Bad data. Not adjustable. Data replaced by FillValue.
5	Value changed	Value changed	Value changed
6	Not used	Not used	Not used
7	Not used	Not used	Not used
8	Interpolated value	Interpolated value	Interpolated value
9	Missing value	Missing value	Missing value

Table A2: Argo Data Management User Manual Table 2.

To address the bias introduced by the removal of levels from the GTS profiles all RTQC are first translated into the form of Table 2a as described above. Profiles in which enough of a disparity in levels exists between the DMQC and RTQC versions of the profile to possibly affect the outcome of a comparison are then excluded. For example, if a profile is flagged as either 'A' or 'F' on Table 2a it will require a disparity in levels of at least 50% to alter the outcome. This is because the intercomparison takes profiles with greater than 50% good levels to be good and fewer than 50% as bad: a profile with 100% good levels will require at least the same number of bad levels to change it's profile rating. Conversely, a profile flagged as 'C' or 'D' requires a disparity of only a single level to call the result into question. The test is performed for each centre and profiles that fail for any centre are excluded from consideration. This process removes 5,939 temperature profiles and 2,658 salinity profiles from the 2010-2011 intercomparison.